Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products

DRAFT GUIDANCE - NOT FOR IMPLEMENTATION

This guidance document is being distributed for comment purposes only.

Draft released for comment on March 13, 1997.

Comments and suggestions regarding this draft document should be submitted by May 30, 1997, to Dockets Management Branch (HFA-305), Food and Drug Administration, 12420 Parklawn Dr., rm. 1-23, Rockville, MD 20857. All comments should be identified with the docket number 97D-0100. For questions regarding this draft document contact Joseph Griffin, (301) 594-5400.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
March 1997

GUIDANCE FOR INDUSTRY[1]
PROVIDING CLINICAL EVIDENCE OF EFFECTIVENESS FOR HUMAN DRUG AND BIOLOGICAL PRODUCTS

I. INTRODUCTION

This document is intended to provide guidance to applicants planning to file NDA's, BLA's, or efficacy supplements on what evidence should be provided to demonstrate effectiveness. Thirty-five years ago, Congress amended the Federal Food, Drug, and Cosmetic Act to add a requirement that manufacturers demonstrate the effectiveness of their products through the conduct of adequate and well-controlled studies to obtain marketing approval. Since then, the issue of what constitutes evidence of effectiveness has been debated by the agency, the scientific community, industry, and others. Sound evidence of effectiveness is a crucial component of the agency's benefit-risk assessment of a new product or use. At the same time, the demonstration of effectiveness represents a major component of drug development time and cost such that the amount and nature of the evidence needed can be an important determinant of when and whether new therapies become available to the public. Thus, the public health is best served by the development of sound evidence of effectiveness in an efficient manner.

The science and practice of drug development and clinical evaluation have evolved significantly since the effectiveness requirement for drugs was established, and this evolution has implications for the amount and type of data needed to support effectiveness in certain cases. As a result of medical advances in the understanding of pathogenesis and disease staging, it is increasingly likely that clinical studies of drugs will be more narrowly defined to focus, for example, on a more specific disease stage or clinically distinct sub-population. As a consequence, product indications are often narrower, the universe of possible indications is larger, and there may be data available from a number of studies of a drug in closely related indications that bear on a determination of its

---

[1]This guidance document has been prepared by the Supplemental Indications Working Group, an agency working group headed by the Deputy Commissioner for Operations. This guidance document represents the agency's current thinking on providing clinical evidence of effectiveness for human drug and biological products. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both. For additional copies of this guidance, contact (1) the Drug Information Branch, Division of Communications Management, HFD-210, CDER, FDA, 5600 Fishers Lane, Rockville, MD 20857 (Phone: 301-827-4573), or (2) the Office of Communication, Training and Manufacturers Assistance, HFM-40, CBER, FDA, 1401 Rockville Pike, Rockville, MD 20852-1448. Send one self-addressed adhesive label to assist that office in processing your requests. The document may also be obtained by mail by calling the CBER Voice Information System at 1-800-835-4709 or 301-827-1800, or by fax by calling the CBER FAX Information System at 1-888-CBERFAX or 301-827-3844. An electronic version of this guidance is also available via Internet using the World Wide Web (WWW). To access the document on the WWW, connect to (1) the CDER Home Page at WWW.FDA.GOV/CDER and go to the "Regulatory Guidance" section or (2) CBER at http://www.FDA.gov/CBER/cberftp.html.

effectiveness for a new use. Similarly, there may be studies of a drug in different populations, studies of a drug alone or in combination, and studies of different doses and dosage forms, all of which may support a particular new use of a drug. At the same time, progress in clinical evaluation has resulted in more rigorously designed and conducted clinical efficacy trials, which are ordinarily conducted at more than one clinical site. This added rigor and scope has implications for a study's reliability, generalizability, and capacity to substantiate effectiveness.

Given this evolution, the agency has determined that it would be appropriate to articulate its current thinking concerning the quantitative and qualitative standards for demonstrating effectiveness of drugs and biologics. FDA hopes that this guidance will enable sponsors to plan drug development programs that are sufficient to establish effectiveness without being excessive in scope and bring greater consistency and predictability to FDA's assessment of the clinical trial data needed to support drug effectiveness. This guidance is also an element of FDA's "*New Use Initiative — Evidence for Primary and Supplemental Approvals*." One goal of this initiative is to encourage submission of supplemental applications to add new uses to the labeling of approved drugs. By articulating how it currently views the quantity and quality of evidence necessary to support approval of a new use of a drug, FDA hopes to illustrate that the submission of supplements for new uses need not be unduly burdensome.

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS

A. Legal Standards

<u>Drugs:</u> The effectiveness requirement for drug approval was added to the Federal Food, Drug, and Cosmetic Act (the Act) in 1962. Between passage of the Act in 1938 and the 1962 amendments, drug manufacturers were required to show only that their drugs were safe. The original impetus for the effectiveness requirement was Congress's growing concern about the misleading and unsupported claims being made by pharmaceutical companies about their drug products, and an environment of high drug prices. After two years of hearings on these issues, Congress adopted the 1962 Drug Amendments, including a provision that required manufacturers of drug products to establish their effectiveness by "substantial evidence." Substantial evidence was defined in section 505(d) of the Act as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."

Since this provision was added to the statute, discussions have ensued regarding the quantity and quality of the evidence needed to establish effectiveness. With regard to quantity, it has been FDA's position that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness. (See e.g., Final Decision on Benylin, 44 FR 51512, 518 (August 31, 1979); Warner-Lambert Co. V. Heckler, 787 F. 2d 147 (3d Cir. 1986)). This position is based on the language in the statute[2] and the legislative history of the 1962 amendments. Language in a Senate report suggested that the phrase "adequate and well-controlled investigations" was designed not only to describe the quality of the required data but the "quantum" of required evidence. S. Rep. No. 1744, Part 2, 87th Cong. 2d Sess. 6 (1962).

Nevertheless, FDA has been flexible within the limits imposed by the Congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug are strong. First, in some cases FDA has relied on pertinent information from other adequate and well-controlled studies of a drug, such as studies of other doses and regimens, of other dosage forms, in other stages of disease, in other populations, and of different endpoints, to support a single adequate and well-controlled study demonstrating effectiveness of a new use. In these cases, although there is only one study of the exact new use, there are in fact multiple studies supporting the new use and expert judgment could conclude that the studies together represent substantial evidence of effectiveness. Second, in some cases FDA has relied on only a single adequate and well-controlled efficacy study to support approval -- generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.

Biologics. Biological products are approved under authority of section 351 of the Public Health Service Act (42 U.S.C.§ 262). Licenses for biologics are to be issued only upon a showing that the products meet standards designed to insure the “continued safety, purity, and potency” of the products (42 U.S.C. § 262 (d)). “Potency” has long been interpreted to include effectiveness (21 CFR 600.3(s)). In 1972, FDA initiated a review of the safety and effectiveness of all previously licensed biologics. The agency stated then that proof of effectiveness would consist of controlled clinical investigations as defined in the provision for “adequate and well-controlled studies” for new drugs (21 CFR 314.126), unless waived as not applicable to the biological product or essential to the validity of the study when an alternative method is adequate to substantiate effectiveness (21 CFR 601.25 (d) (2)). One such adequate alternative was identified to be serological response data where a previously accepted correlation with clinical effectiveness exists. As with

[2] Section 505(d) of the Act uses the plural form in defining “substantial evidence” as “adequate and well-controlled investigations, including clinical investigations.” See also use of “investigations” in section 505(b) of the Act, which lists the contents of a new drug application.

non-biological drug products, FDA has approved biological products based on single, multicenter studies with strong results.

B. Scientific Basis for the Legal Standard

The requirement for more than one adequate and well-controlled investigation reflects the need for *independent substantiation* of experimental results. A single clinical experimental finding of effectiveness, unsupported by other independent evidence, has not normally been considered adequate to support a conclusion of effectiveness. The reasons for this include the following:

- Any clinical trial may be subject to unanticipated, undetected, systematic biases. These biases may operate despite the best intentions of sponsors and investigators, and may lead to flawed conclusions. In addition, some investigators may bring conscious biases to evaluations.

- The inherent variability in biological systems may produce a positive trial result by chance alone. This possibility is acknowledged (and quantified to some extent) in the statistical evaluation of most efficacy results. Independent substantiation of a favorable result protects against the possibility that a chance occurrence will lead to an erroneous conclusion that a treatment is effective.

- Results obtained in a single center may be dependent on site or investigator specific factors (e.g., disease definition, concomitant treatment, diet). In such cases, the results, although correct, may not be generalizable to the intended population. This possibility is the primary basis for the importance of independence in substantiating studies.

- Rarely, efficacy results are the product of scientific fraud.

Independent substantiation of experimental results addresses these problems by providing consistency across more than one study, thus greatly reducing the possibility of a biased, chance, site-specific, or fraudulent result.

The need for independent substantiation has often been referred to as the need for "replication" of the finding. Replication may not be the best term, however, as it may imply that precise repetition of the same experiment in other patients by other investigators is the only means to substantiate a conclusion. Precise replication of a trial is only one of a number of possible means of obtaining independent substantiation of a clinical finding and, at times, can be less than optimal as it could leave the conclusions vulnerable to any systematic biases inherent to the particular study design. Results that are obtained from studies that are independent in both design and execution, perhaps evaluating different populations, endpoints, or dosage forms, may provide support for a conclusion of effectiveness that is as convincing as, or more convincing than, a repeat of the same study.

C. Guidance Re: The Quantity of Evidence to Support Effectiveness

The following three sections provide guidance on the quantity of evidence needed in particular circumstances to establish substantial evidence of effectiveness. Section 1 addresses situations in which effectiveness of a new use may be extrapolated entirely from existing efficacy studies. Section 2 addresses situations in which there is other information from adequate and well-controlled studies that supports the findings of a controlled trial of a specific new use, such as studies in other phases of a disease, closely related diseases, other conditions of use (different dose, duration of use, regimen), different dosage forms, or different endpoints. Section 3 addresses situations in which typically, a single, multicenter study may provide evidence that a use is effective.

In each of these situations, it is assumed that any studies relied upon to support effectiveness meet the requirements for adequate and well-controlled studies in 21 CFR 314.126. It should also be appreciated that reliance on a single study of a given use, whether alone or with substantiation from related trial data, leaves little room for study imperfections or contradictory (non-supportive) information. In all cases, it is presumed that the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post-hoc changes in analysis, or other factors is judged to be minimal, and that the results reflect a clear prior hypothesis documented in the protocol. Moreover, a single favorable study among several similar attempts that failed to show effectiveness would not ordinarily constitute persuasive support for a product use.

This guidance is not intended to be a complete listing of the circumstances in which existing effectiveness data may provide independent substantiation of related claims; rather, it provides examples of the reasoning that may be employed. The examples are applicable whether the claim arises in the original filing of an NDA or BLA, or in a supplemental application.

1. Extrapolation from Existing Studies

In certain cases, effectiveness of an approved drug product for a new indication may be adequately demonstrated without additional adequate and well-controlled efficacy trials. Ordinarily, this will be because other types of data provide a way to apply the known effectiveness to a new population or dosage form. The following are examples of situations in which effectiveness can be extrapolated from efficacy data for another claim or product.

a. Pediatric uses

The rule revising the Pediatric Use section of product labeling (21 CFR 201.57(f)(9)) makes allowance for inclusion of pediatric use information in labeling without controlled clinical trials of the use in children. In such cases, a sponsor must provide other information to support pediatric use and the agency must conclude that the course of the disease and the effects of the drug are sufficiently similar in the pediatric and adult populations to permit extrapolation from adult efficacy data to pediatric patients. Evidence that could support a conclusion of similar disease course and similar drug effect in adult and pediatric populations includes evidence of common pathophysiology and natural history of the disease in adults and in the pediatric population, evidence of common drug metabolism and mechanism of action in each population, and experience with the drug, or other drugs in its therapeutic class, in the disease or condition or related diseases or conditions. Examples in which pediatric use labeling information has been extrapolated from adult efficacy data include ibuprofen and ondansetron for chemotherapy-induced nausea and vomiting.

b. Bioequivalence

Studies of alternative formulations may rely on evaluation of bioequivalence. New dosage strengths are routinely accepted based on bioavailability.

c. Pharmacokinetic linking

In some cases, modified release dosage forms may be approved on the basis of pharmacokinetic data linking the new dosage form to a previously studied immediate-release dosage form. Because the pharmacokinetic patterns of controlled-release and immediate-release dosage forms are not identical, it is generally important to have some understanding of the relationship of blood concentration to response to extrapolate to the new dosage form.

2. Demonstration of Effectiveness by a Single Study of a New Use, with Independent Substantiation From Related Existing Study Data

The discussion that follows describes specific examples in which a single study of a new use, with independent substantiation from existing study data in related uses, could provide evidence of effectiveness. In these cases, the study in the new use and the related studies constitute the “adequate and well-controlled investigations” that support the conclusion that the drug has the effect it is purported to have. Whether related studies are capable of substantiating a single study of a new use is a matter of judgment.

a. Different doses, regimens, or dosage forms

Dose-response relationships are generally continuous such that information about the effectiveness of one dose, dosage regimen, or dosage form is relevant to the effectiveness of other doses, regimens, or dosage forms as long as blood levels and exposure are not very different. In cases where the pharmacokinetics of a new regimen are very similar to the previously approved one, no further clinical trial support may be needed. In this situation, pharmacokinetic data is used to bridge the results of the new dose, regimen or dosage form to the efficacy trial results (See section II.C.1.b). In cases where the pharmacokinetics are not so similar, clinical data will likely be necessary to support effectiveness of a new regimen, but a single additional efficacy study should ordinarily be sufficient.

b. Studies in other phases of the disease

In many cases, therapies that are effective in one phase of a disease are effective in other disease phases, although the magnitude of the benefit and benefit-to-risk relationship may be different in other disease phases. For example, if a drug is known to be effective in patients with a refractory stage of a particular cancer, a single adequate and well-controlled study of the drug in an earlier stage of the same tumor will generally be sufficient evidence of effectiveness to support the new use.

c. Studies in other populations

Often, responses in subsets of a particular patient population are qualitatively similar to those in the whole population when the course of disease and effect of the drug are sufficiently similar. In most cases, for a drug already shown to be generally effective in a condition or to be effective in one population, a single study would be sufficient to support effectiveness in age, race, gender, concomitant disease, or other subsets.

d. Studies in combination or as monotherapy

For a drug known to be effective as monotherapy, a single adequate and well-controlled study usually is sufficient to support effectiveness of the drug when combined with other therapy. Similarly, known effectiveness of a drug as part of a combination (*i.e.*, its contribution to the effect of the combination is known) would allow a single study to support its use as monotherapy, or as part of a new combination, for the same use. For example, a single study designed to demonstrate statistically comparable immune response will ordinarily provide sufficient efficacy data for a new combination vaccine containing products or antigens already proven to be effective alone or in other combinations. Also, single studies each showing effectiveness of a drug as part of different combinations would support claims for the drug's effectiveness in both combinations. These situations are common for oncologic and antihypertensive drugs.

e. Studies in a closely related disease

Studies in etiologically or pathophysiologically related conditions, or studies of a symptom common to several diseases (e.g., pain) can support each other, allowing initial approval of several uses or allowing additional claims based on a single adequate and well-controlled study. For example, certain anticoagulant therapies could be approved for use in two different settings based on individual studies in unstable angina/acute coronary syndrome and in the postangioplasty state. Because the endpoints studied and the theoretical basis for use of an anti-coagulant are suitably similar, each study supports the other for each claim. Similarly, single analgesic studies in several painful conditions would ordinarily be sufficient to support either a general claim or multiple specific claims.

f. Studies in less closely related diseases, but where the general purpose of therapy is similar

Certain classes of drug therapy, such as antimicrobials and antineoplastics, are appropriate interventions across a range of different diseases. For therapies of this type, evidence of effectiveness in one disease could provide independent substantiation of effectiveness in a quite different disease. For example, it is possible to argue that effectiveness of an antibiotic in any infectious disease setting may support reliance on a single study showing effectiveness in other settings where the organism is sensitive, the drug is present in sufficient concentrations, and the disease pathophysiologies and patient populations are generally similar.[3] Similarly, for an oncologic drug, effectiveness in one or more tumor types may support reliance on a single study showing effectiveness against a different kind of tumor, especially if the tumor types have a common biological origin.

g. Studies of different clinical endpoints

Demonstration of a beneficial effect in different studies on two different clinically meaningful endpoints could cross-substantiate a claim for effectiveness for each outcome. For example, the initial claim for effectiveness of enalapril for heart failure was supported by one study showing symptom improvement over several months and a second study showing improved survival in a more severely ill population. The two different findings, each from an adequate and well-controlled study, led to the conclusion that enalapril was effective in both treating symptoms and improving survival.

h. Pharmacologic/pathophysiologic endpoints

When the pathophysiology of a disease and the pharmacodynamics of a therapy are very well understood, it may be possible to link specific pharmacologic effects to a strong likelihood of clinical effectiveness. In some cases, these pharmacologic effects are accepted as surrogate endpoints, either supporting ordinary approval (*e.g.*, blood pressure effects, cholesterol-lowering effects) or accelerated approval under 21 CFR 314 Subpart H and 21 CFR 601 Subpart E (*e.g.*, CD4 count or viral load effects to support effectiveness of drugs for HIV infection). When the

[3] See Division of Anti-Infective Drug Products: Points to Consider in the Clinical Development and Labeling of Anti-Infective Drug Products, October 1992.

pharmacologic effect is not sufficient alone as an effectiveness endpoint, but the linkage between it and the clinical outcome is strong, not merely on theoretical grounds but based on prior therapeutic experience or well-understood pathophysiology, a single adequate and well-controlled study showing clinical efficacy can be substantiated by persuasive data showing the related pharmacologic effect.

For example, a single clearly positive trial can be sufficient to support approval of a replacement therapy such as a coagulation factor, when it is combined with clear evidence that the condition being treated is caused by a deficiency of that factor. Demonstration of physical replacement of the deficient factor or restoration of the missing physiologic activity provides strong substantiation of the clinical effect. The corrective treatment of an inborn error of metabolism could be viewed similarly. In the case of preventive vaccines, one adequate and well-controlled clinical trial may be supported by compelling animal challenge/protection models, human serological data, passive antibody data, or pathogenesis information. The more evidence there is linking effects on the physiologic endpoint to improvement or prevention of the disease, the more persuasive the argument for reliance on one study.

Note, however, that plausible pharmacologic effects have often not correlated with clinical benefit, and, therefore, caution must be observed in relying on a pharmacologic effect as evidence of effectiveness. For example, pharmacologic effects such as arrhythmia suppression by Type 1 antiarrhythmics and increased cardiac output by phosphodiesterase inhibitor or beta adrenergic inotropes resulted in increased mortality rather than, as was expected, decreased sudden death and improved outcome in heart failure. Generally, the utility of pharmacologic outcomes in providing independent substantiation will be greatest where there is prior experience with the pharmacologic class.

i. Efficacy of *in vivo* diagnostic products

There are a number of situations in which a single adequate and well-controlled trial of an *in vivo* diagnostic product, when combined with additional information about the product, might provide sufficient evidence to demonstrate its effectiveness.

(1) As with therapeutic products, information from other doses, regimens, or formulations of a diagnostic product, or in other

populations or phases of disease, and, in the case of imaging agents, other imaging protocols might be used to substantiate the results of a single new study of the product for a new indication. For example, a single study demonstrating efficacy of a diagnostic product as a confirmatory test in patients in whom an alternative test is positive might be substantiated by a study of the diagnostic performance of the new test in patients in whom the alternative test was negative or equivocal. Also, information about drug pharmacology, such as the extent of drug uptake or distribution in biochemical or metabolic processes or known information about receptor binding in a specific disorder, might indicate it would be appropriate to extend the results of a single diagnostic study to a related disorder.

(2) Where the efficacy of a class of *in vivo* diagnostics is well established and can be reasonably applied to future agents in the class (*e.g.*, the iodinated contrast agents), a single multicenter study would ordinarily be sufficient to establish efficacy for a new agent in the class. In this case, the indication or indications for the product would be identical to the well established indications of the class. However, safety of the new product and its relationship to dose and other pertinent factors would need to be evaluated in an adequate number of patients.

(3) During clinical evaluation, diagnostic products often undergo a series of trials, each of which is designed to evaluate a different aspect of the diagnostic test. For example, often a diagnostic product will be tested in a series of patients already known to be positive or negative to establish the optimal cutoff value or image characteristics for the diagnostic's desired sensitivity and specificity. This procedure ordinarily demonstrates the ability of the test to perform as a diagnostic. The diagnostic is then evaluated in a series of "unknown" patients to assess whether the desired sensitivity, specificity, and predictive values are attained in a patient population similar to that expected during clinical use. In the typical trial in unknowns, the actual status of each test patient is ascertained in several ways *(e.g.*, multiple diagnostic tests, clinical evaluation, biopsy, clinical follow-up) to confirm the performance of the diagnostic. In this type of diagnostic development program, substantiating data may be contributed both by the earlier trial evaluating diagnostic performance in known patients, and by the

multiple confirmatory diagnostic tests performed during the efficacy trial in unknown patients. If these data show a strong separation between true positive and false positive results, a single, multicenter efficacy trial may be sufficient to confirm diagnostic efficacy.

Although a second efficacy trial may not be needed to substantiate effectiveness in these situations, this does not mean that other important characteristics of effectiveness evidence, such as independence and generalizability, can be neglected. For example, estimates of the effectiveness of certain diagnostics, such as imaging agents, may vary among clinical sites or populations, and thus may be dependent on the training of the test readers or on characteristics of the patients studied. In these cases, it is crucial that the role of factors such as inter-reader variability and population differences be rigorously assessed.

3. Evidence of Effectiveness from a Single Study

Thirty-five years ago, when the effectiveness requirement was originally implemented, the prevailing study model was a single institution, single investigator, relatively small trial with relatively loose blinding procedures, and little attention to prospective identification of outcomes and analyses. At present, major clinical efficacy studies are typically multicentered, with clear, prospectively determined clinical and statistical analytic criteria. These studies are less vulnerable to certain biases, are often more generalizable, may achieve extreme statistical results, and can often be evaluated for consistency across subgroups, centers, and multiple endpoints.

The added rigor, power, and scope of contemporary clinical trials have made it possible to rely, in certain circumstances, on a single adequate and well-controlled study, without independent substantiation from another controlled trial, as a sufficient scientific and legal basis for approval. For example, the approval of timolol for reduction of post-infarction mortality was based on a particularly persuasive (low p-value), internally consistent, single multicenter study that demonstrated a major effect on mortality and reinfarction rate. For ethical reasons, the study was considered unrepeatable. The Center for Biologics Evaluation and Research has also approved a number of products based upon a single persuasive study. The agency provided a general statement in 1995 describing when a single, multicenter study may suffice (60 FR 39181; August 1, 1995), but the agency has not comprehensively described the situations in which a single, adequate and well-controlled study alone might be considered to support an effectiveness claim, or the characteristics of a single study that could make it adequate support for an effectiveness claim.

It should be appreciated that relying on a single adequate and well-controlled study is inevitably a matter of judgment and that the conclusion based on even a highly persuasive single study will be less secure than a conclusion based on two similar studies. For this reason, reliance on a single study will generally be limited to situations in which a trial has demonstrated a clinically meaningful effect on mortality, irreversible morbidity, or prevention of a disease with potentially serious outcome, such that confirmation of the result in a second trial would be ethically difficult or impossible. For example, sequential repetition of long-term, strongly positive trials that demonstrated a decrease in post-infarction mortality, prevention of osteoporosis, or prevention of pertussis could present significant ethical concerns.

The discussion that follows identifies the characteristics of a single adequate and well-controlled study that could make the study adequate support for an effectiveness claim. While no one of these characteristics is necessarily determinative, the presence of one or more in a study can contribute to a conclusion that the study would be adequate to support an efficacy claim.

a. Large Multicenter Study

In a large multicenter study in which (1) no single study site provided an unusually large fraction of the patients and (2) no single investigator or site was disproportionately responsible for the effect seen, concerns about lack of generalizability of the finding or an inexplicable result involving a single investigator are lessened. If analysis shows that a single site is largely responsible for the effect, however, the credibility of a multicenter study is diminished.

b. Multiple "studies" in a single study

Large multicenter studies may have prospective stratifications or identified analytic subsets based on variables such as disease severity, geographic residence, or demographic characteristics. Where the strata are randomized separately and each shows a significant effect, the study provides two or more separate estimates of the effect, albeit not by independent investigators and often not with a clear prospective intent to do so.

Properly designed factorial studies may be analyzed as a series of pairwise comparisons, representing separate demonstrations of activity of a drug as monotherapy and in combination with another drug. This model was successfully used in ISIS II, which showed that for patients with a myocardial infarction both aspirin and streptokinase had favorable effects on survival when used alone and when combined (aspirin alone and streptokinase alone were each superior to placebo; aspirin and streptokinase in combination were superior to aspirin alone and to streptokinase alone). This represented two separate (except that investigators and sites were the same) demonstrations of the effectiveness of aspirin and streptokinase.

c. Multiple endpoints involving different events

In some cases, a single study will include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect. Where a study shows statistically persuasive evidence of an effect on more than one of such endpoints, the internal weight of evidence of the study is enhanced. For example, the approval of beta-interferon (Betaseron) for prevention of exacerbations in multiple sclerosis was based on a single multicenter study, at least partly because there was both a decreased rate of exacerbations and a decrease in MRI-demonstrated disease activity -- two entirely different, but logically related, endpoints. Similarly, favorable effects on both death and non-fatal myocardial infarctions in a lipid-lowering, post-angioplasty, or post-infarction study would, in effect, represent different, but consistent, demonstrations of effectiveness, greatly reducing the possibility that a finding of reduced mortality was a chance occurrence. In contrast, a beneficial effect on multiple endpoints that evaluate essentially the same phenomenon and correlate strongly, such as mood change on two different depression scales or SGOT and CPK levels post-infarction, does not significantly enhance the internal weight of the evidence from a single trial.

Although two consistent findings within a single study provide reassurance that neither finding is due to chance, they do not protect against selection bias or biased analyses. For example, a treatment assignment not well balanced for important prognostic variables could lead to an apparent effect on both endpoints. Thus, close scrutiny of study design and conduct are critical to evaluating this type of study.

d. Statistically very powerful finding

In a multicenter study, an extreme p-value indicates that the result is highly inconsistent with the null hypothesis of no treatment effect. In some studies it is possible to detect nominally significant results in data from several centers, but even where that is not possible an overall extreme result and significance level means that most study centers had similar findings. For example, the thrombolysis trials of streptokinase (ISIS II, GISSI) had very sizable treatment effects and extreme p-values, greatly adding to their persuasiveness. Preventive vaccines for infectious disease indications with a high efficacy rate (*e.g.*, point estimate of efficacy of 80% or higher and a reasonably narrow 95% confidence interval) have been approved based on a single adequate and well-controlled trial.

4. Reliance on a Single, Multicenter Study — Caveats

Although acknowledging the persuasiveness of a single, internally consistent, "strong" multicenter study, it must be appreciated that there remains a possibility that even a strong result can represent an isolated or biased result. Recently, the apparent highly favorable effect of vesnarinone, an inotropic agent, in heart failure (60% reduction of mortality in what appeared to be a well-designed, placebo-controlled, multicenter trial with an extreme p-value) has proven to be unrepeatable. In an attempt to substantiate the finding, the same dose of the drug that seemed lifesaving in the earlier study significantly increased mortality (by 26%), and a lower dose also appeared to have a detrimental effect on survival.

When considering whether to rely on a single multicenter trial, it is critical that the possibility of an incorrect outcome be considered and that all the available data be examined for its potential to either support or undercut reliance on a single multicenter trial. In the case of vesnarinone, there were other data that were not consistent with the dramatically favorable outcome in the multicenter study. These data seemed to show an inverse dose-response relationship, showed no suggestion of symptomatic benefit, and showed no effect on hemodynamic endpoints. These inconsistencies led the agency, with the advice of its Cardio-Renal Advisory Committee, to refuse approval -- a decision borne out by the results of the subsequent study.

The case of tirilizad, a drug intended to improve survival of patients with subarachnoid hemorrhage, is also instructive. An initial European study, which enrolled mostly women, demonstrated an overall increase in survival ($p<0.01$), but with the entire effect observed in men. The result in women had no favorable trend. A second study, requested by the agency, found no effect at all in men or women. Although explanations for the odd initial result and the negative second study were offered (low blood levels in women and in U.S. patients receiving phenytoin), none has been convincingly documented and the agency did not accept the single study as sufficient to demonstrate effectiveness.

These examples illustrate that inadequacies and inconsistencies in the data, such as lack of pharmacologic rationale, failure of related drugs to show the observed effect, and lack of expected "other" effects accompanying a critical outcome, can weaken the persuasiveness of a single trial. In addition, although an unexplained failure to substantiate the results of a favorable study in a second controlled trial is not proof that the favorable study was in error, it is a strong reason not to rely on the single favorable study.

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFICACY CLAIM

When submitting the requisite quantity of data to support approval of a new product or new use of an approved product, sponsors must also document that the studies were adequately designed and conducted. Thus, the quality of scientific evidence to support effectiveness is of comparable importance to the quantity of evidence. Essential characteristics of adequate and well-controlled trials are described in 21 CFR 314.126. In order to demonstrate that a trial supporting an efficacy claim is adequate and well-controlled, extensive documentation of trial planning, protocols, conduct, and data handling is usually submitted to the agency, and detailed patient records are made available at the clinical sites.

From a scientific standpoint, however, it is recognized that the extent of documentation necessary depends on the particular study, the types of data involved, and the other evidence available to support the claim. Therefore, the agency is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be assured. This section discusses the factors that influence the extent of documentation needed, with particular emphasis on studies evaluating new uses of approved drugs.

For the purposes of this section, the phrase "documentation of the quality of evidence" refers to (1) the completeness of the documentation and (2) the ability to access the primary study data and the original study related records (e.g., subjects' medical records, drug accountability records) for the purposes of verifying the data submitted as evidence.
These interrelated elements bear on a determination of whether a study is adequate and well-controlled. In practice, to achieve a high level of documentation, studies supporting claims are ordinarily conducted in accordance with Good Clinical Practices (GCP's). Sponsors routinely monitor all clinical sites, and FDA routinely has access to the original clinical protocols, primary data, clinical site source documents for on-site audits, and complete study reports.

However, there are often situations in which studies that evaluate the effectiveness of a drug product lack the full documentation described above (for example, full patient records may not be available) or in which the study was conducted with less monitoring than is ordinarily seen in commercially sponsored trials. These situations are more common for supplemental indications because post-approval studies are more likely to be conducted by parties other than the drug sponsor and those parties may employ less extensive monitoring and data gathering procedures than a sponsor. Notwithstanding less than usual documentation or monitoring, under certain circumstances it is possible for sponsors to rely on such studies to support effectiveness claims. Some of those circumstances are described below.

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports

FDA's access to primary data has proven to be important in many regulatory decisions. FDA also has reason to be skeptical of the conclusions of published reports of studies. Experience has shown that such study reports do not always contain a complete, or entirely accurate, representation of study plans and outcomes. Outright fraud (i.e., deliberate deception) is unusual. However, incompleteness, lack of clarity, unmentioned deviation from planned analyses, or an inadequate description of how critical endpoint judgments or assessments were made are common flaws. Typically, journal article peer reviewers only have access to a limited data set and analyses, and thus may lack sufficient information to detect critical gaps and problems. The utility of peer review can also be affected by variability in the relevant experience and expertise of peer reviewers. FDA's experiences with the Anturane Reinfarction Trial, as well as literature reports of the effectiveness of tacrine and the anti-sepsis HA-1A antibody illustrate its concerns with reliance on the published medical literature.

Notwithstanding these concerns, the presence of some of the factors discussed below can make it possible to rely on studies for which FDA has less than usual access to data or detailed study reports to partially or entirely (the so-called "paper" filing) support an efficacy claim. There is a distinction between relying entirely on a published report, and relying on a published report supported by additional important information about the study. Reliance on a literature report to support an efficacy claim is more likely if FDA can obtain additional critical study details. Section 1 below describes additional information that, if available, would increase the likelihood that a study could be relied upon to support an effectiveness claim. Section 2 describes factors that may make a finding of effectiveness sufficiently persuasive to permit reliance on the published literature alone. Note that the factors outlined in Section 2 are relevant to an assessment of the reliability of literature reports generally, whether alone or accompanied by other important information as discussed in Section 1.

1. Submission of Published Literature or Other Reports in Conjunction with Other Important Information that Enhances the Reliability of the Data

If a sponsor wishes to rely on a study conducted by another party and cannot obtain the primary data from the study, for most well-conducted studies it will be possible to obtain other important information, such as a protocol documenting the prospective plans for the trial, records of trial conduct and procedures, patient data listings for important variables, and documentation of the statistical analysis. FDA has considerable experience in evaluating large multicenter outcome studies sponsored by U.S. and European government agencies (NIH, British Medical Research Council) and private organizations (the ISIS studies, the SAVE study) for which there was limited access to primary study data, but for which other critical information was available. Providing as many as possible of the following important pieces of information about a study, in conjunction with the published report, can increase the likelihood that the study can be relied upon to support an efficacy claim:

a. The protocol used for the study, as well as any important protocol amendments that were implemented during the study and their relation to study accrual or randomization.

b. The prospective statistical analysis plan and any changes from the original plan that occurred during or after the study, with particular note of which analyses were performed pre- and post-unblinding.

c. Randomization codes and documented study entry dates for the subjects.

d. Full accounting of all study subjects, including identification of any subjects with on-treatment data who have been omitted from analysis, and an analysis of results using all subjects with on-study data.

e. Electronic or paper record of each subject's data for critical variables and pertinent baseline characteristics. Where individual subject responses are a critical variable (*e.g*., objective responses in cancer patients, clinical cures and microbial eradications in infectious disease patients, death from a particular cause), detailed bases for the assessment, such as the case report, hospital records, and narratives, should be provided when possible.

f. Where safety is a major issue, complete information for all deaths and drop-outs due to toxicity. For post-approval supplemental uses, however, there is generally less need for the results of lab tests or for details of adverse event reports and, consequently, much more limited documentation may be sufficient (*e.g.*, only for unexpected deaths and previously undescribed serious adverse effects). Exceptions to this approach would include situations in which the population for the supplemental use is sufficiently different that existing safety information has limited application (*e.g.*, thrombolysis in stroke patients versus myocardial infarction patients) or where the new population presents serious safety concerns (e.g., extension of a preventive vaccine indication from young children to infants).

2. Submission of Published Literature Reports Alone

The following factors increase the possibility of reliance on published reports alone to support approval of a new product use:

a. Multiple studies conducted by different investigators where each of the studies clearly has an adequate design and where the findings across studies are consistent.

b. A high level of detail in the published reports, including clear and adequate descriptions of statistical plans, analytic methods (prospectively determined), and study endpoints, and a full accounting of all enrolled patients.

c. Clearly appropriate endpoints that can be objectively assessed and are not dependent on investigator judgment (e.g., overall mortality, blood pressure, or microbial eradication). Such endpoints are more readily interpreted than more subjective endpoints such as cause-specific mortality or relief of symptoms.

d. Robust results achieved by protocol-specified analyses that yield a consistent conclusion of efficacy and do not require selected post hoc analyses such as covariate adjustment, subsetting, or reduced data sets (e.g., analysis of only responders or compliant patients, or of an "eligible" subset).

e. Conduct of studies by groups with properly documented operating procedures and a history of implementing such procedures effectively.

There have been approvals based primarily or exclusively on published reports. Examples include the initial approval of secretin for evaluation of pancreatic function and recent approvals of bleomycin for malignant pleural effusion and doxycycline for malaria.

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring

Industry-sponsored studies typically use extensive on-site and central monitoring and auditing procedures to assure data quality. Studies supported by other sponsors may employ less stringent procedures and may use no on-site monitoring at all. An International Conference on Harmonization guideline on Good Clinical Practices, recently accepted internationally, emphasizes that the extent of monitoring in a trial should be based on trial-specific factors (e.g., design, complexity, size, and type of study outcome measures), and that different degrees of outside monitoring can be appropriate. In recent years, many credible and valuable studies conducted by government or independent study groups, often with important mortality outcomes, had very little on-site monitoring. These studies have addressed quality control in other ways such as by close control and review of documentation, and extensive guidance and planning efforts with investigators. There is a long history of reliance on such studies for primary as well as supplementary applications. Factors that influence whether studies with limited or no monitoring may be relied on include the following:

1. The existence of a prospective plan for the assurance of data quality.

2. Studies that have features that make them inherently less susceptible to bias, such as those with relatively simple procedures, non-critical entry criteria, and readily assessed outcomes.

3. The ability to sample critical data and make comparisons to supporting records (e.g., hospital records).

4. Conduct of the study by a group with established operating procedures and a history of implementing such procedures effectively.